Evidence Retrieval for Fact Verification

Abhibhu Prakash **Anirban Haldar** **Saikat Moi** **Pranil Dey**

ABSTRACT

For the aim of fact-checking in this project, we use the FEVER dataset to obtain documents in accordance with the search. There are a total of 145449 claims in the FEVER Dataset. The assertions are divided into 'SUPPORTED', 'REFUTED', and 'NOTENOUGHINFO' categories. We are left with 109810 claims on elimination of claims marked 'NOTENOUGHINFO'. The query and the claims are tokenized from the list of documents, and cosine similarity is used to rank them. We next validate the fact from the obtained documents using the LSTM model.

I. PROBLEM STATEMENT

Fact or claim verification is a two-step process. First, you retrieve supporting or refuting evidence related to a claim. Then based on the set of evidence snippets, the task is to determine whether the claim is true or false. In this project, we are mainly focused on the first step, i.e., evidence retrieval.

II. MOTIVATION

The volume of online data has increased at least as swiftly as the computer's speed. The propagation of false information is becoming a serious issue as a result of the rise of diverse online news sources and social media, endangering public safety and having negative societal, political, and economic effects. Internet use allows people and organizations to spread potentially misleading information to a large audience, whether on purpose or accidentally. Such false information frequently distorts reality, incites unwarranted fear, and sows discord. For instance, while the U.S. presidential election was taking place, dozens of youngsters in the Macedonian town of Veles generated false information on social media and made a lot of money. This is an example of the allurement of power and profit associated with fake news. Fact-checking is now necessary due to the accessibility of ever-increasing amounts of information mixed with the ease of sharing over the internet. While receiving a lot of attention in the journalism world, fact checking is crucial for various fields and applications, including news, science, product evaluations, and the verification of medical claims. Fact checking manually requires a lot of time and mental effort, which is why it is only done by qualified experts. Several websites, such as HoaxSlayer, PolitiFact, FactCheck and others, have recently developed to offer expert-based fact-verification. To determine if a piece of news is accurate or not, PolitiFact uses three editors. Because manual verification cannot handle the vast amount of false material on the internet and cannot respond rapidly, it is only useful to a limited extent. As a result, there has been a high need for automation to be made available in order to lessen the time and workload required of humans while undertaking fact checking. Finding textual evidence, followed by evidence reasoning and entailment, is a difficult and complex task in automated fact checking. Reasoning offers a scientific theory to explain why the assertion is supported by the data. In natural language processing, involvement is the direction of a relationship between text fragments.Researchers have taken notice of this automatic fact verification. With the aim of finding inaccurate or misleading information, various computational approaches have been put forth to automate verification tasks. Existing methods typically use traditional sparse retrieval, which can have poor memory, especially when the pertinent passages have little to no overlap, to retrieve pertinent evidence from a collection of documents. Many jobs just take into account how an article's body and headlines interact; they do not take into account the requirement to uncover relevant evidence, as a human would. Because the amount of information that needs to be reviewed grows exponentially, there is always a need to work towards developing an efficient automated solution to combat online falsehoods.

III. RELATED WORK

With the explosion in information (and consequently misinformation), our need for fact verification has also increased and advancements in evidence retrieval and fact verification are trying to keep up with this demand. Fake news propagation has become a serious

menace. Traditionally there have been 2 primary ways to tackle this.
One of them is framing it as a classification problem. However this is not always feasible as in real life, much of the news we come across cannot be strictly assigned as either true or false as they are often partially accurate and partially not. We can assign separate additional classes for this type of news. The outputs are multi-class labels which can be learned as independent tables when we use these datasets. Since gathering accurate labels is labor intensive, therefore we often use unsupervised or semi supervised methods to achieve the task.
Another common approach is to convert it into a regression problem. Here the result is a numeric score which signifies the truthfulness. Here we generally obtain the score by taking the difference of the predicted score and the ground truth score or by taking Pearson/ Spearman correlations. The difficulty encountered in this method is that the available datasets have discrete ground truth scores and hence it is hard to convert these labels into numeric scores. To overcome this, we mostly use supervised methods.
Here, we have experimented with different approaches which include tf-idf, bm25, word2vec. The details are described in the upcoming sections.

IV. TECHNIQUE AND EXPERIMENTS

We have experimented with different approaches like tf-idf, bm25, and word2vec models using this dataset to retrieve evidence for a given query.

1. tf-idf: This is a method for generating numerical representations of text documents by assigning weights to the words based on their frequency and rarity. We have first pre-processed the claims and after that, we have created a tf-idf matrix for them. We have tokenized and pre-processed the query text and created a tf-idf vector for it. After that, we have calculated the cosine similarity between the query vector and each document vector in the tf-idf matrix. After that, we have printed the top 5 most similar pieces of evidence for the given query.

2. bm25: It is a probabilistic model, which means that it produces scores that can be interpreted as probabilities of relevance. This makes it more suitable for information retrieval. After the preprocessing part, we have created the bm25 index for the preprocessed claims. We have also tokenized and preprocessed the query text. After that, we calculated the bm25 scores. We have printed the top 5 pieces of evidence for the given query.

3. Word2Vec: This is a neural network based method for generating dense vector representations of words in a large corpus.
For evidence extraction we first used the Word2Vec model of the 'gensim' library and then used cosine similarity.
To implement this, we had to use stemming and tokenization. We tokenized the claims using the package 'nltk'. While tokenizing the claim sentences, we also removed the stop-words and punctuations.

Query: Nikolaj Coster-Waldau
Output: ['Gujarat', 'Saratoga_-LRB-film-RRB-', 'Superman', 'Justin Trudeau', 'Kris_Wu']

LSTM: After the documents are retrieved, we use the LSTM model to verify the query. The model was run on the query. The model was run on the test data, and we got the following results.

Labels	Precision	Recall	f1-score
negative	0.66	0.36	0.46
positive	0.80	0.94	0.86

To further enhance the performance of this model, we increased the number of LSTM units in the first layer to 128 and added a second layer with 64 units. We also decreased the learning rate.

We ran 3 queries on the enhanced model. We got positive (supports) for the first 2 queries and negative (refutes) for the last query which are correct output. The accuracy is found to be 0.78.

Query: ["Slovenia uses the euro.", "Saratoga is an American film from 1937.", "Massachusetts is far away from Rhode island."]

Output:

positive
positive
negative

V. ANALYSIS AND FUTURE THOUGHTS

Since the time individuals have started exchanging information, fact checking has been a constant trouble. Due to the original data being extremely dimensional, noisy, and sparse, cosine similarity performed poorly. The dearth of training samples led to a dip in the performance of LSTM on documents with label refutation. To categorize the claims as supporting or disputing, we used the LSTM model.

WORK DISTRIBUTION

Name	Roll No.	Work Done
Abhibhu Prakash	20CS10002	LSTM, pre-processing, ppt
Anirban Haldar	20CS10008	Word2vec, report
Saikat Moi	20CS10050	tf-idf, bm25, report
Pranil Dey	20CS30038	LSTM, pre-processing, ppt

REFERENCES

1. Christopher Manning, Prabhakar Raghavan, Hinrich Schutze, “Introduction to Information Retrieval”, Cambridge University Press
2. Jie Zhou, Xu Han, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li and Maosong Sun, “GEAR: Graph-based Evidence Aggregating and Reasoning for Fact Verification” Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, , August 2019, pp. 892–901
3. L. Graves, “Understanding the Promise and Limits of Automated Fact Checking”, Technical Report, Reuters Institute, University of Oxford, February 2018.
4. Shyam Subramanian and Kyumin Lee, “Hierarchical Evidence Set Modeling for Automated Fact Extraction and Verification”, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), November 2020, pp. 7798-7809
5. Giannis Bekoulis, Christina Papagiannopoulou and Nikos Deligiannis, (In press) “A Review on Fact Extraction and Verification”, ACM Computing Surveys, Volume 55, Issue 1, January 2023, pp. 1–35
6. Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Tuo Zhao, “SMART: Robust and Efficient Fine-Tuning for Pre-trained Natural Language Models through Principled Regularized Optimization”, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, Association for Computational Linguistics, July 2020, pp. 2177–2190
7. Chris Samarinas, Wynne Hsu, and Mong Li Lee, “Improving Evidence Retrieval for Automated Explainable Fact-Checking”, Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations, Online, Association for Computational Linguistics, June 2021, pp. 84–91.
8. Bernhard Kratzwald and Stefan Feuerriegel, “Adaptive Document Retrieval for Deep Question Answering”, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, November 2018, pp. 576–581
9. .James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal, “FEVER: a Large-scale Dataset for Fact Extraction and verification”, Proceedings of NAACL-HLT, 2018, pp. 809–819
10. Fan Yang, Eduard Dragut, Arjun Mukherjee, “Improving Evidence Retrieval with Claim-Evidence Entailment”, Proceedings of Recent Advances in Natural Language Processing, September 2021, pp.1553–1558